Artificial Intelligence

Linear Regression

Pham Viet Cuong
Dept. Control Engineering & Automation, FEEE
Ho Chi Minh City University of Technology

Linear Regression

What is regression analysis?

- ✓ A basic and commonly used predictive analysis (predictive analysis: the use of data, statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data)

Examples: Sales manager trying to predict next month's numbers

- ✓ Monthly sales numbers for, e.g., the past three years
- ✓ Weather
- ✓ Competitor's promotion
- ✓ Rumor of a new and improved model
- ✓ . . .

Linear Regression

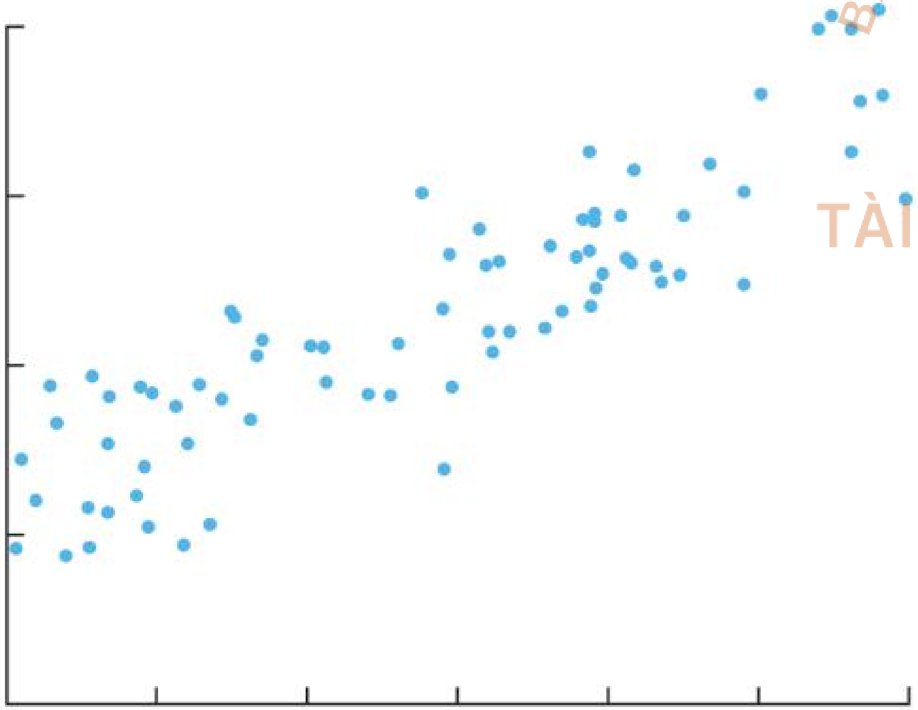
Regression analysis: examine relationship between one dependent variable and one or more independent variables.

- ✓ Can a set of predictor variables predict an outcome (dependent) variable?
- ✓ Which variables matter most?
- ✓ Which can we ignore?
- ✓ How do those factors interact with each other?
- ✓ How certain are we about all of these factors?

Linear Regression

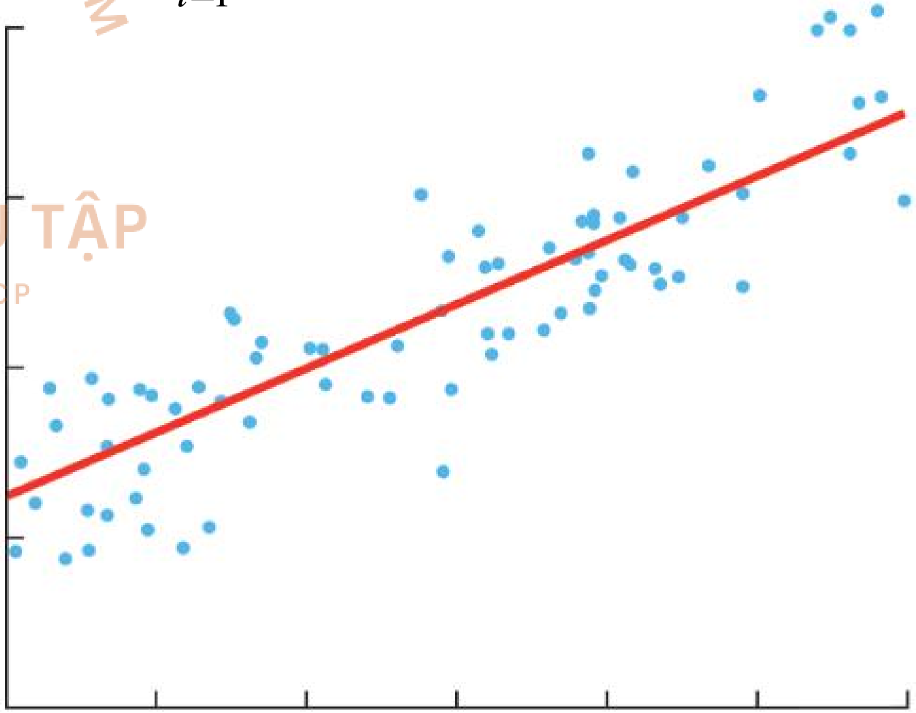
Simple linear regression

$$f(x) = w_0 + w_1 x$$



Least-square linear regression problem

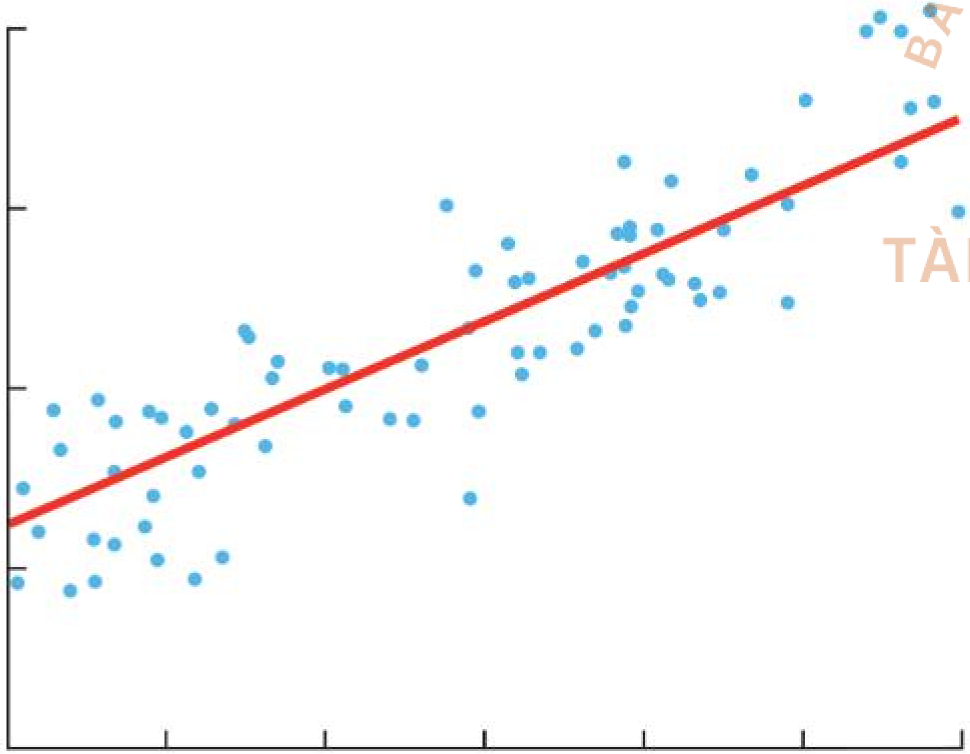
$$\min_{w_0, w_1} \sum_{i=1}^{N} \left[y_i - \left(w_0 + w_1 x_i\right)\right]^2$$



Linear Regression

Simple linear regression

$$f(x) = w_0 + w_1 x$$



Least-square linear regression problem

$$\min_{w_0, w_1} \frac{1}{2} \sum_{i=1}^{N} \left[y_i - (w_0 + w_1 x_i) \right]^2$$

$$w_1^* = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i \right)^2}$$

$$w_0^* = \bar{y} - w_1^* \bar{x}$$

Linear Regression

Multiple Linear Regression

$$f(\mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_p x_p$$

$$\mathbf{w} = \left[w_0, w_1, w_2, \ldots, w_p\right]^T$$

$$\bar{\mathbf{x}} = \left[1, x_1, x_2, \ldots, x_p\right]$$

$$f(\mathbf{x}) = \bar{\mathbf{x}}\mathbf{w}$$

Loss function

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}(y_i - \bar{\mathbf{x}}_{\mathbf{i}}\mathbf{w})^2$$

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \mathcal{L}(\mathbf{w})$$

Linear Regression

$$\mathbf{y} = [y_1; y_2; \ldots; y_N]$$

$$\bar{\mathbf{X}} = [\bar{\mathbf{x}}_1; \bar{\mathbf{x}}_2; \ldots; \bar{\mathbf{x}}_N]$$

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}(y_i - \bar{\mathbf{x}}_i\mathbf{w})^2$$

$$= \frac{1}{2}\|\mathbf{y} - \bar{\mathbf{X}}\mathbf{w}\|_2^2 \quad (3)$$

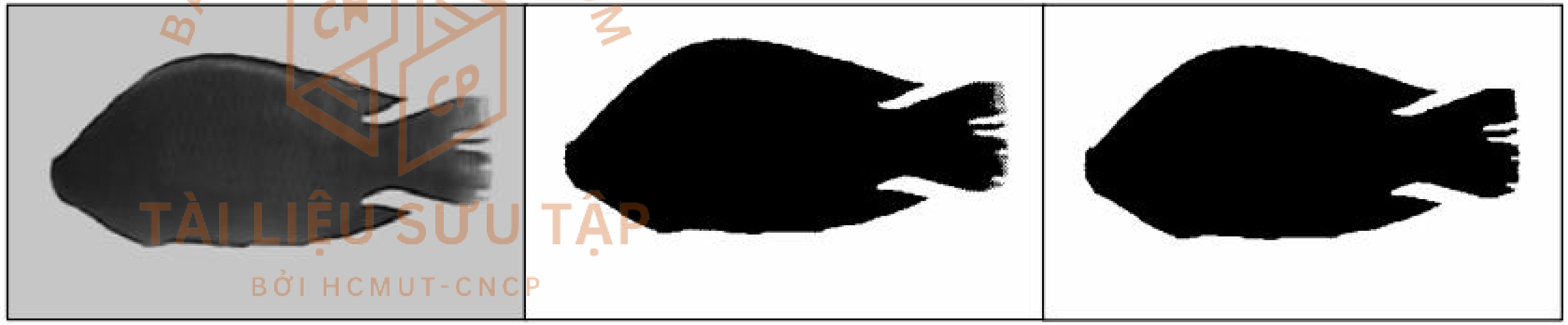
$$\frac{\partial \mathcal{L}(\mathbf{w})}{\partial \mathbf{w}} = \bar{\mathbf{X}}^T(\bar{\mathbf{X}}\mathbf{w} - \mathbf{y})$$

$$\mathbf{w} = \mathbf{A}^\dagger\mathbf{b} = (\bar{\mathbf{X}}^T\bar{\mathbf{X}})^\dagger\bar{\mathbf{X}}^T\mathbf{y}$$

$$\|\boldsymbol{x}\|_2 := \sqrt{x_1^2 + \cdots + x_n^2}$$

Linear Regression

Example: Simple Linear Regression

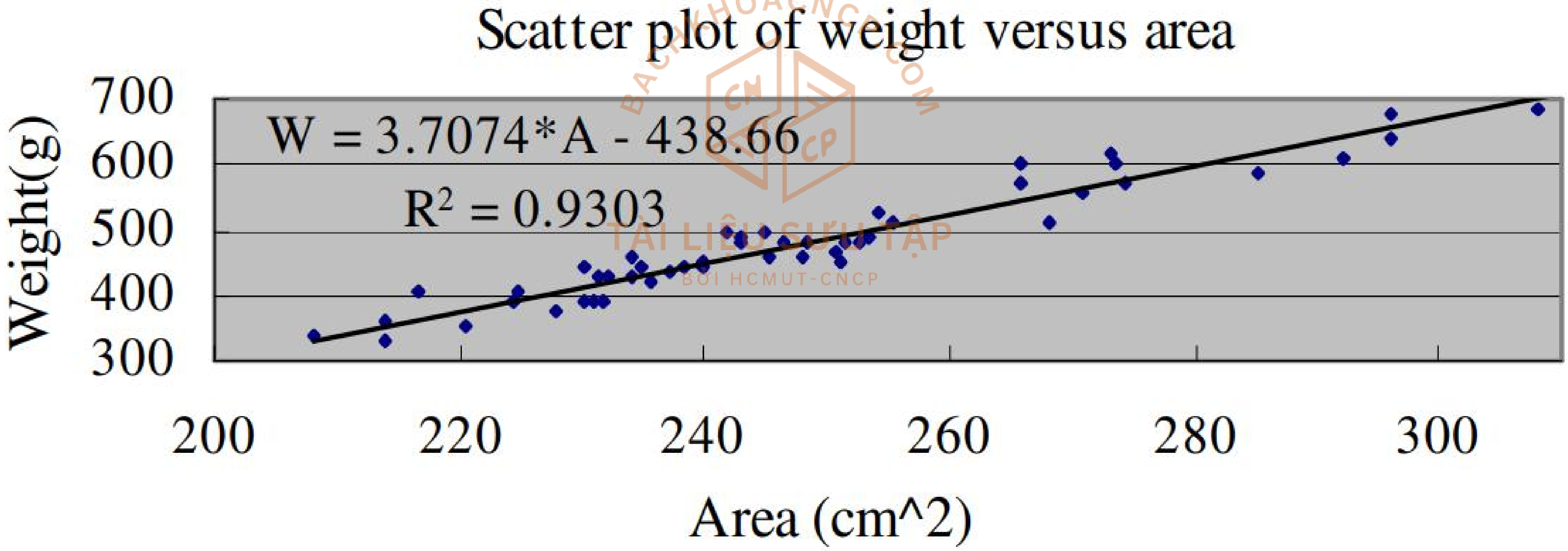


Linear Regression

No.	W(g)	A(cm^2)	No.	W(g)	A(cm^2)	No.	W(g)	A(cm^2)	No.	W(g)	A(cm^2)
1	388	230.30	14	570	274.25	27	490	242.90	40	638	296.00
2	428	234.06	15	492	253.58	28	480	252.60	41	602	273.70
3	352	220.60	16	496	244.89	29	512	255.20	42	600	266.00
4	374	227.70	17	462	248.10	30	460	233.91	43	616	273.30
5	450	251.07	18	436	237.16	31	484	248.40	44	584	285.30
6	462	245.40	19	432	231.26	32	528	254.20	45	482	251.35
7	332	214.10	20	464	250.66	33	498	2420	46	454	240.08
8	338	208.20	21	418	235.76	34	442	239.9	47	484	242.99
9	444	238.35	22	442	234.70	35	610	292.10	48	484	246.30
10	388	231.8	23	408	224.60	36	746	309.80	49	440	230.10
11	408	216.77	24	428	232.06	37	678	296.20	50	388	224.29
12	388	231.08	25	558	270.70	38	574	265.95			
13	360	214.10	26	512	268.30	39	684	308.10			

Linear Regression

$$W = 3.7074 * A + 438.66$$



Linear Regression

Max error: 9.12g (1.97%)
Mean error: 1.77g (0.83%)
Standard deviation: 2.21g

No.	A (cm^2)	W_1(g)	W_2(g)	ε (g)	ε %	No.	A (cm^2)	W_1(g)	W_2(g)	ε (g)	ε %
1	238.18	442	444.35	-2.35	0.53	26	229.86	408	413.52	-5.52	1.35
2	228.93	408	410.06	-2.06	0.50	27	223.53	388	390.04	-2.04	0.53
3	232.81	428	424.46	3.54	0.83	28	213.97	360	354.62	5.38	1.49
4	268.24	558	555.82	2.18	0.39	29	223.66	388	390.52	-2.52	0.65
5	256.07	512	510.70	1.30	0.25	30	233.38	428	426.57	1.43	0.33
6	244.83	462	469.02	-7.02	1.52	31	214.35	352	356.02	-4.02	1.14
7	208.04	332	332.62	-0.62	0.19	32	271.03	570	566.15	3.85	0.68
8	209.42	338	337.75	0.25	0.07	33	250.95	492	491.72	0.28	0.06
9	237.50	444	441.84	2.16	0.49	34	302.08	678	681.26	-3.26	0.48
10	223.90	388	391.42	-3.42	0.88	35	272.21	574	570.52	3.48	0.61
11	251.62	490	494.21	-4.21	0.86	36	219.83	374	376.33	-2.33	0.62
12	248.92	480	484.19	-4.19	0.87	37	248.18	484	481.46	2.54	0.52
13	254.25	512	503.96	8.04	1.57	38	236.43	440	437.89	2.11	0.48
14	275.26	584	581.84	2.16	0.37	39	222.62	388	386.67	1.33	0.34
15	248.89	482	484.07	-2.07	0.43	40	245.94	464	473.12	-9.12	1.97
16	242.18	454	459.19	-5.19	1.14	41	233.04	418	425.33	-7.33	1.75
17	250.58	484	490.35	-6.35	1.31	42	240.95	460	454.65	5.35	1.16
18	254.09	498	503.35	-5.35	1.07	43	247.60	484	479.28	4.72	0.98
19	236.20	442	437.04	4.96	1.12	44	262.96	528	536.24	-8.24	1.56
20	284.78	610	617.12	-7.12	1.17	45	279.96	602	599.27	2.73	0.45
21	320.09	746	748.06	-2.06	0.28	46	281.14	600	603.65	-3.65	0.61
22	254.47	496	504.77	-8.77	1.77	47	285.70	616	620.53	-4.53	0.74
23	241.99	462	458.48	3.52	0.76	48	304.12	684	688.83	-4.83	0.71
24	236.70	436	438.89	-2.89	0.66	49	289.38	638	634.19	3.81	0.60
25	233.59	432	427.35	4.65	1.08	50	238.07	450	443.95	6.05	1.34

Linear Regression

Sources:

1. Yu-Teng Liang and Yih-Chih Chiou. "Machine Vision-Based Automatic Raw Fish Handling and Weighing System of Taiwan Tilapia." B.-C. Chien et al. (Eds.): IEA/AIE 2009, LNAI 5579, pp. 711–720, 2009
2. http://www.mit.edu/~6.s085/notes/lecture3.pdf
3. https://www.statisticssolutions.com/what-is-linear-regression/
4. https://hbr.org/2015/11/a-refresher-on-regression-analysis
5. https://www.surveygizmo.com/resources/blog/regression-analysis/
6. https://machinelearningcoban.com/2016/12/28/linearregression/

Linear Regression

No.	W(g)	A(cm^2)	No.	W(g)	A(cm^2)	No.	W(g)	A(cm^2)	No.	W(g)	A(cm^2)
1	388	230.30	14	570	274.25	27	490	242.90	40	638	296.00
2	428	234.06	15	492	253.58	28	480	252.60	41	602	273.70
3	352	220.60	16	496	244.89	29	512	255.20	42	600	266.00
4	374	227.70	17	462	248.10	30	460	233.91	43	616	273.30
5	450	251.07	18	436	237.16	31	484	248.40	44	584	285.30
6	462	245.40	19	432	231.26	32	528	254.20	45	482	251.35
7	332	214.10	20	464	250.66	33	498	2420	46	454	240.08
8	338	208.20	21	418	235.76	34	442	239.9	47	484	242.99
9	444	238.35	22	442	234.70	35	610	292.10	48	484	246.30
10	388	231.8	23	408	224.60	36	746	309.80	49	440	230.10
11	408	216.77	24	428	232.06	37	678	296.20	50	388	224.29
12	388	231.08	25	558	270.70	38	574	265.95			
13	360	214.10	26	512	268.30	39	684	308.10			

Linear Regression

fish_data1 - Notepad

File Edit Format View Help

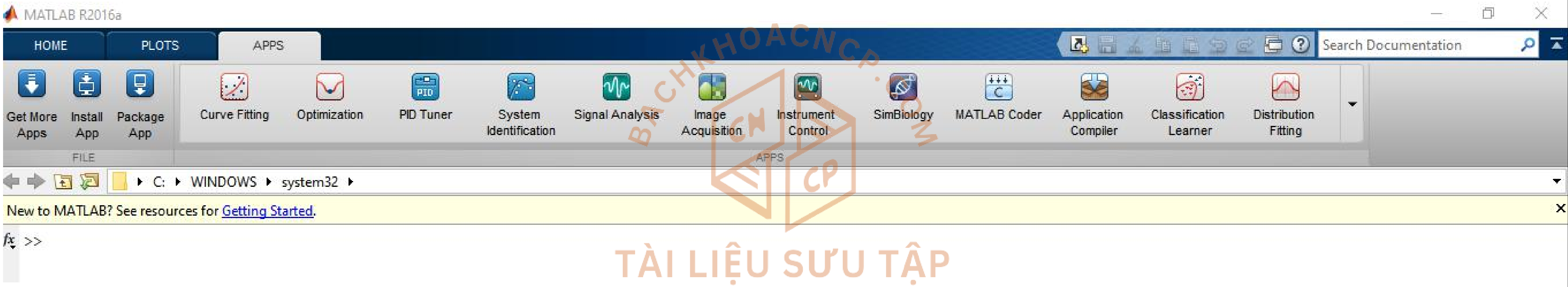
1 388 230.30 14 570 274.25 27 490 242.90 40 638 296.00 2 428 234.06 15 492 253.58 28 480 252.60 41 602 273.70 3 352 220.60 16 496 244.89 29 512 255.20 42 600 266.00 4 374 227.70 17 462 248.10 30 4

Linear Regression

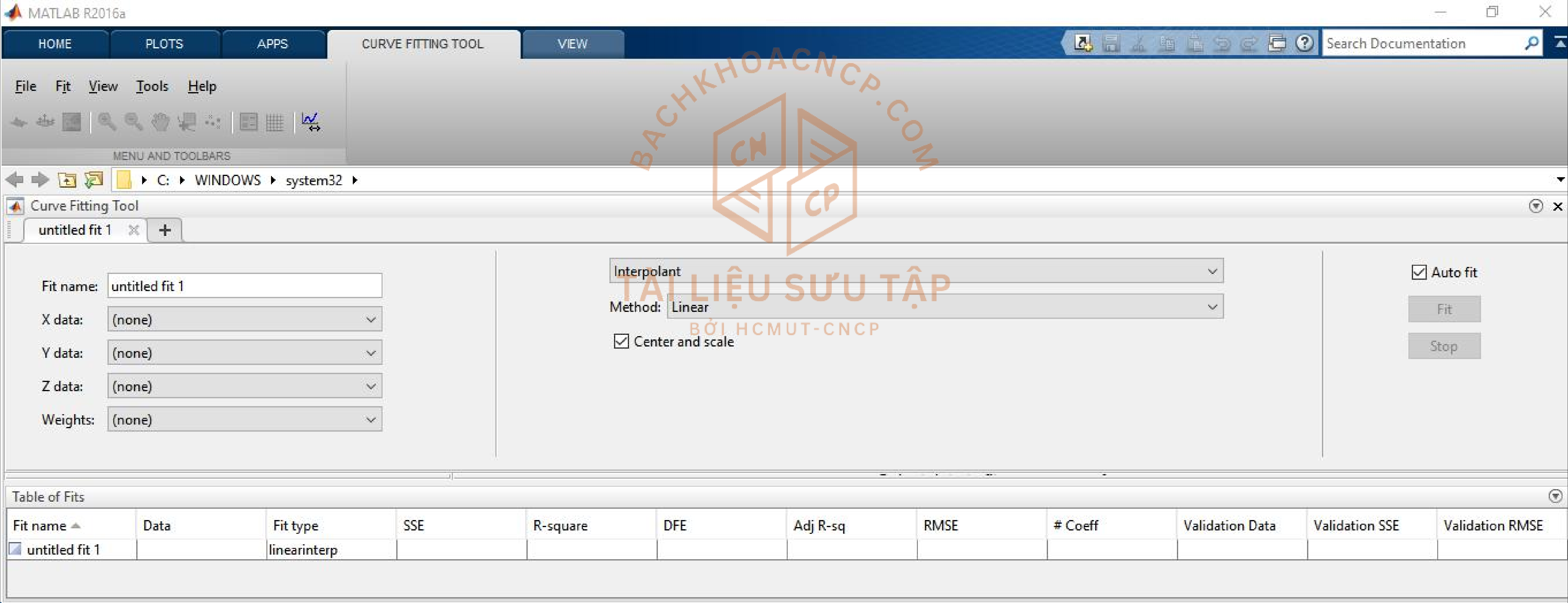
read_data.m

```
clear
clc

fileID = fopen('D:\Research\Shrimp\fish_data1.txt');
C1 = textscan(fileID,'%f');
data1 = C1{1,1};
fclose(fileID);

fileID = fopen('D:\Research\Shrimp\fish_data2.txt');
C2 = textscan(fileID,'%f');
data2 = C2{1,1};
fclose(fileID);

W1 = data1(2:3:end);
A1 = data1(3:3:end);

A2 = data2(2:6:end);
W2 = data2(3:6:end);
W2_bar = data2(4:6:end);
```

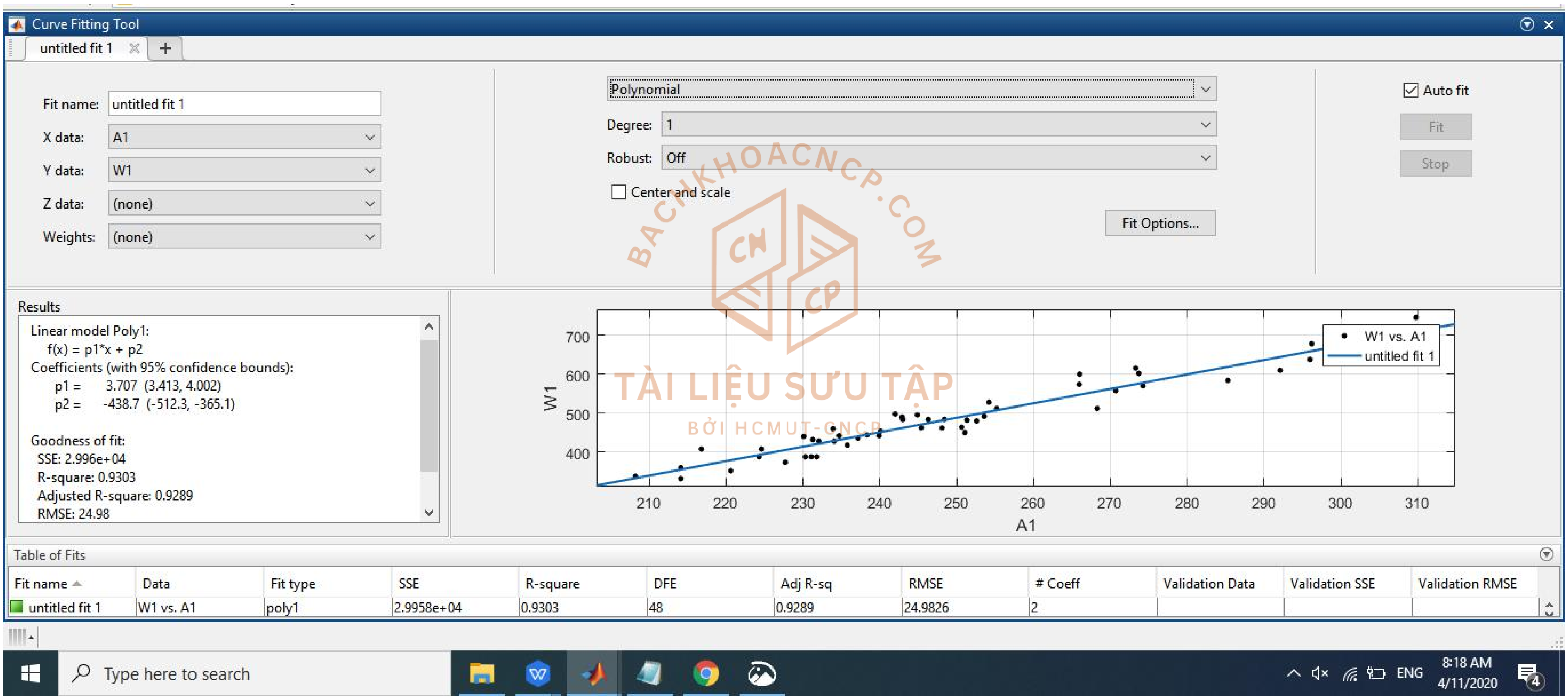
Linear Regression



Linear Regression



Linear Regression



Linear Regression

Linear Regression

Linear Regression

Linear Regression